Adaptive estimation on anisotropic Hölder spaces Part II. Partially adaptive case

Nicolas Klutchnikoff

Adaptive estimation on anisotropic Hölder spaces
Part II. Partially adaptive case

Nicolas Klutchnikoff*

March 18, 2006

Abstract

In this paper, we consider a particular case of adaptation. Let us recall that, in the first paper "Fully case", a large collecton of anisotropic Hölder spaces is fixed and the goal is to construct an adaptive estimator with respect to the absolutely unknown smoothness parameter. Here the problem is quite different: an additionnal information is known, the *effective smoothness* of the signal. We prove a minimax result which demonstrates that a knowledge of is type is useful because the rate of convergence is better than that obtained without knowledge of the effective smothness. Moreover we linked this problem with the maxiset theory.

1 Introduction

1.1 Statistical model

This paper is the second part of our paper "Fully adaptive case". Further, we will refer to this paper as (Part I). We consider the same model. Our observations $\mathcal{X}^{(\varepsilon)} = (X_\varepsilon(u))_{u\in[0,1]^d}$ satisfies the same SDE:

$$X_\varepsilon(du) = f(u)du + \varepsilon W(du), \quad \forall u \in [0,1]^d,$$

where $f : \mathbf{R}^d \to \mathbf{R}$ is an unknown signal to be estimated, W is a standartd Gaussian white noise from $\mathbf{R}^d$ to $\mathbf{R}$ and ε is the noise level.

Our main goal is to estimate f at a fixed point $t \in (0,1)^d$.

*Université de Provence, LATP, UMR CNRS 6632. Mail: `klutchni@cmi.univ-mrs.fr`

1.2 Our goal

In this second part of our article, we study the "Partially adaptive case". Let us recall that we are interrested in pointwise estimation among the class of anisotropic Hölder spaces. Let us recall some notations: $l_* < l^*$ and $b = (b_1, \dots, b_d)$ are given. Moreover, we consider only Hölder spaces $H(\beta, L)$ (defined in Part I) such that

$$\beta \in \mathcal{B} = \prod_{i=1}^{d} (0; b_i] \text{ and } L \in \mathcal{I} = [l_*; l^*].$$

Remark 1. *Let us just recall that $\beta = (\beta_1, \dots, \beta_d)$ can be viewed as the smoothness parameter. Each β_i represents the smoothness of a function in direction i. Moreover, L is a Lipschitz constant.*

We denote $\Sigma = \bigcup_{\mathcal{B}\times\mathcal{I}} H(\beta, L)$. Our goal is to answer this questions: Is it possible to guarantee a quality of estimation? On which space (included in Σ)? With which procedure of estimation?

For example, if we consider $\tilde{\eta}_\varepsilon(\gamma) = \varepsilon^{2\gamma/(2\gamma+1)}$, it is well known that we can guarantee this quality on each space $H(\beta, L)$ such that $\bar{\beta} = \gamma$ (because it is the minimax rate of convergence on this space) using the minimax on this space estimator. But one of our results implies that we cannot guarantee this quality simultaneousely on each such space.

Now, we fix $0 < \gamma < \bar{b}$, and we consider

$$\eta_\varepsilon(\gamma) = (l^*)^{\frac{1}{2\gamma+1}} \left(\|K\| \varepsilon \sqrt{\ln \ln \frac{1}{\varepsilon}} \right)^{\frac{2\gamma}{2\gamma+1}}.$$

Our result is that there exists an estimator, namely $f_\varepsilon^\gamma(\cdot)$, such that $\eta_\varepsilon(\gamma)$ is the minimax rate of convergence of this estimator on $\Sigma(\gamma)$ defined by

$$\Sigma(\gamma) = \bigcup_{(\beta,L)\in\mathcal{B}(\gamma)\times\mathcal{I}} H(\beta, L) = \bigcup_{\beta\in\mathcal{B}(\gamma)} H(\beta, l^*)$$

where

$$\mathcal{B}(\gamma) = \left\{ \beta \in \mathcal{B} : \bar{\beta} = \gamma \right\}.$$

Thus, using f_ε^γ as procedure of estimation, we can guarantee that the quality is $\eta_\varepsilon(\gamma)$, at least on $\Sigma(\gamma)$.

1.3 Result

Theorem 1. *Our result consists in two inequalities:*

$$\limsup_{\varepsilon\to 0} \sup_{f\in\Sigma(\gamma)} \mathbf{E}_f\left[\left(\eta_\varepsilon(\gamma)^{-1}|f_\varepsilon^\gamma(t)-f(t)|\right)^q\right] < +\infty. \tag{U.B.}$$

$$\liminf_{\varepsilon\to 0} \inf_{\tilde f} \sup_{f\in\Sigma(\gamma)} \mathbf{E}_f\left[\left(\eta_\varepsilon(\gamma)^{-1}|\tilde f(t)-f(t)|\right)^q\right] > 0, \tag{L.B.}$$

where the infimum is taken over all possible estimators.

In words, f_ε^γ is a minimax on $\Sigma(\gamma)$ estimator.

This paper consists in the proof of this assertion. First, we construct the estimator f_ε^γ. Next, we prove the corresponding lower bound.

Remark 2. *Let us remark that this result can be viewed as an adaptive result. Indeed, let us consider $\Sigma(\gamma)$ as a family —instead of an union— of Hölder spaces $H(\beta,L)$ such that $\bar\beta=\gamma$. It is well known that on each $H(\beta,L)$ there exists a minimax on this space estimator which depends explicitely on (β,L) at least trough its bandwidth. Thus, question of adaptation arrizes naturally.*

Our lower bound proves that an optimal adaptive estimator $f^*(\cdot)$ such that

$$\limsup_{\varepsilon\to 0} \sup_{(\beta,L)\in\mathcal{B}(\gamma)\times\mathcal{I}} \sup_{f\in H(\beta,L)} \mathbf{E}_f\left[\left(\varepsilon^{-\frac{2\gamma}{2\gamma+1}}|f^*(t)-f(t)|\right)^q\right] < +\infty$$

does not exist.

Our upper bound proves that $f_\varepsilon^\gamma(\cdot)$ is an adaptive estimator. Moreover the price to pay is only $\sqrt{\ln\ln 1/\varepsilon}$ which is to be compared with the classical loss $\sqrt{\ln 1/\varepsilon}$ in other adaptive problems.

Moreover we prove that our estimator is optimal in a minimax sense.

2 Procedure

2.1 Collection of kernel estimators

Let us recall that kernels were defined in the first part of this paper: "fully adaptive case". Here we have just to chose a good collection of kernel estimators.

Let us define

$$n_\varepsilon(\gamma) = \left\lfloor \frac{1}{\ln 2}\left(\frac{4(\bar b-\gamma)}{(2\bar b+1)(2\gamma+1)}\ln\frac{l^*}{\|K\|\varepsilon} - \frac{1}{2\gamma+1}\ln\ln\ln\frac{1}{\varepsilon}\right)\right\rfloor.$$

Let us denote
$$\mathcal{Z}_\gamma^\varepsilon = \mathcal{Z}(n_\varepsilon(\gamma)).$$
Let us recall the definition of this set:
$$\mathcal{Z}(n) = \left\{ k =\in \mathbf{Z}^d : \sum_{i=1}^{d}(k_i+1) = n \text{ and } \forall i, |k_i| \le C(b)n+1 \right\},$$
where
$$C(b) = \frac{2\bar{b}+1}{2\bar{b}} \times \frac{\ln 2 + \sqrt{2\ln 2}}{\ln 2}.$$

Finally, we consider the following collection $\{\hat{f}_k(\cdot)\}_{k\in\mathcal{Z}_\gamma^\varepsilon}$.

2.2 Notations

Let us recall the following notation: for all $k \in \mathcal{Z}_\gamma^\varepsilon$, we have
$$\sigma_\varepsilon(k) = \frac{\varepsilon\|K\|}{\left(\prod_{i=1}^d h_i^{(k)}\right)^{1/2}},$$
where $h^{(k)} = (h_1^{(k)}, \dots, h_d^{(k)})$ is defined by:
$$h_i^{(k)} = (\|K\|\varepsilon)^{\frac{2\bar{b}}{2\bar{b}+1}\frac{1}{b_i}} 2^{-(k_i+1)}.$$
It is clear that for all k and l in $\mathcal{Z}_\gamma^\varepsilon$, $\sigma_\varepsilon(k) = \sigma_\varepsilon(l) \triangleq \sigma_\varepsilon(\gamma)$ and moreover that:
$$\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}} \asymp \eta_\varepsilon(\gamma).$$
Following the same strategy as in the first part of our paper, let us define the set $\mathcal{A}$ as follows: an index $k \in \mathcal{Z}_\gamma^\varepsilon$ belongs to $\mathcal{A}$ if it satisfies:
$$\left|\hat{f}_{k\wedge l}(t) - \hat{f}_l(t)\right| \le C\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}}, \quad \forall l \ne k, l \in \mathcal{Z}_\gamma^\varepsilon,$$
where $k \wedge l$ denote the index $(k_i \wedge l_i)_{i=1,\dots,d}$.

2.3 Definition of our procedure

First of all, let us reformulate one of our result obtained in the first part of this paper: there exists an estimator, namely $f_\varepsilon^\Phi(\cdot)$, sucht that
$$\limsup_{\varepsilon\to 0}\sup_{\beta\in\mathcal{B}}\sup_{f\in H(\beta,l^*)} \mathbf{E}_f\left[\left(\left(\varepsilon\sqrt{\ln\frac{1}{\varepsilon}}\right)^{-\frac{2\bar{\beta}}{2\bar{\beta}+1}} |f_\varepsilon^\Phi(t) - f(t)|\right)^{2q}\right] < +\infty.$$

Now, let us define our new estimator: If the random set $\mathcal{A}$ is non-empty, we chose arbitrary any index which belongs to this set. We denote $\hat{k}$ a such index. Then we construct:

$$f_\varepsilon^\gamma(\cdot) = \hat{f}_{\hat{k}}(\cdot).$$

On the other hand, if $\mathcal{A}$ is empty, we define

$$f_\varepsilon^\gamma(\cdot) = f_\varepsilon^\Phi(\cdot).$$

Remark 3. *This procedure is closed to the adaptive one. The main difference consists in the following: when the set $\mathcal{A}$ is empty, we estimate using a best estimator than* 0. *In fact the probability $\mathbf{P}_f[\mathcal{A} = \emptyset]$ is too large to use a trivial estimator.*

3 Proof of (U.B)

3.1 Method

First of all, let us recall that our minimax on Σ_γ estimator is in fact an "adaptive procedure of estimation" because the real smoothness parameter is unknown.

Thus, the mechanism of the proof is very closed to the previous one. We will compare the estimator chosen by our procedure with respect to the "best" estimator among our class but depending on the unknown parameter.

First, we have to define correctly all indexes we need. Next, we will be able to prove the result. Moreover, as the class Σ_γ depends only in L though l^* (because $H(\beta, L) \subset H(\beta, l^*)$), we will assume that $l_* = l^* = 1$ to make the proof simpler. Consequently we will denote $H(\beta)$ instead of $H(\beta, 1)$.

3.2 Indexes

Let us suppose that our unknown signal in Σ_γ belongs to $H(\beta)$ with $\bar{\beta} = \gamma$. Clearly, if we consider the kernel estimator defined using bandwidth

$$\left(h_i(\beta, \varepsilon) = \left(\|K\|\varepsilon\sqrt{\ln\ln\frac{1}{\varepsilon}} \right)^{\frac{2\gamma}{2\gamma+1}\frac{1}{\beta_i}} \right)_{i=1,\dots,d},$$

it achieves the expected rate $\eta_\varepsilon(\gamma)$.

We consider the bandwidth

$$h^*(\varepsilon) = \left(h_i^*(\varepsilon) = (\|K\|\varepsilon)^{\frac{2\bar{b}}{2\bar{b}+1}\frac{1}{b_i}} \right)_{i=1,\dots,d}$$

and define the following indexes: for all $i \in [\![1;d]\!]$ and $\beta \in \mathcal{B}$ such that $\bar{b} = \gamma$, we construct

$$\tilde{k}_i(\beta,\varepsilon) = \left\lfloor \frac{1}{\ln 2} \ln \frac{h_i^*(\varepsilon)}{h_i(\beta,\varepsilon)} \right\rfloor .$$

If $h^{(k)} = (h_i^{(k)})_i$ denote the bandwidth defined by

$$h_i^{(k)} = h_i^*(\varepsilon) 2^{-(k_i+1)},$$

we obtain clearly that the kernel estimator defined using bandwidth $h^{(\tilde{k}(\beta,\varepsilon))}$ is asymptotically as good as that one defined using $h(\beta,\varepsilon)$.

Now, let us define:

$$k_i(\beta,\varepsilon) = \begin{cases} \tilde{k}_i(\beta,\varepsilon) & \text{if } i = 1,\dots,d-1 \\ n_\varepsilon(\gamma) - 1 - \sum_{i=1}^{d_1} (\tilde{k}_i(\beta,\varepsilon)+1) & \text{otherwise} \end{cases}$$

It is easy to prove that

$$\left| \tilde{k}_d(\beta,\varepsilon) - k_d(\beta,\varepsilon) \right| \leq d.$$

Thus, asymptotically, estimator defined using $h^{(k(\beta,\varepsilon))}$ is as good as that one defined using $h^{(\tilde{k}(\beta,\varepsilon))}$ and thus as good as that one defined by $h(\beta,\varepsilon)$.

Moreover it is simple, by producing similar arguments than in the first part of this paper, to obtain that $k(\beta,\varepsilon)$ belongs to $\mathcal{Z}(n_\varepsilon(\gamma))$.

3.3 Proof

We want to prove that, for all $\varepsilon < 1$:

$$\sup_{\beta \in \mathcal{B}(\gamma)} \sup_{f \in H(\beta)} \mathbf{E}_f \left[\left(\eta_\varepsilon^{-1}(\gamma) \left| f_\varepsilon^\gamma(t) - f(t) \right| \right)^q \right] < M_q(\gamma)$$

where $M_q(\gamma)$ is an explicit constant given in the proof.

Set $\varepsilon < 1$ and $\beta \in \mathcal{B}$ such that $\bar{\beta} = \gamma$. Let us suppose that $f \in H(\beta) \subset \Sigma_\gamma$. Set q a fixed parameter.

First, let us suppose that $\mathcal{A}$ is non empty.

A) $\mathcal{A}$ is non empty

Let us denote $\kappa = k(\beta,\varepsilon)$. Our goal is to majorate the following quatiy:

$$(*) = \mathbf{E}_f \left[|\hat{f}_{\hat{k}}(t) - f(t)|^q \right].$$

Let us consider

$$\begin{cases} I_1 &= |\hat{f}_{\hat{k}}(t) - \hat{f}_{\hat{k}\wedge\kappa}| \\ I_2 &= |\hat{f}_{\hat{k}\wedge\kappa}(t) - \hat{f}_\kappa(t)| \\ I_3 &= |\hat{f}_\kappa(t) - f(t)| \end{cases}$$

Let us remark that, if $\hat{k} = \kappa$, then $I_1 = I_2 = 0$. Thus we can suppose that $\hat{k} \neq \kappa$.
a) Let us control of $\mathbf{E}_f[I_3^q]$. Using lemma **??**, we have:

$$\begin{array}{rcl} \mathbf{E}_f[I_3^q] & = & \mathbf{E}_f\left[|\hat{f}_\kappa(t) - f(t)|^q\right] \\ & \leq & \mathbf{E}_f[(|b_\kappa(t) - f(t)| + \sigma_\varepsilon(\gamma)|\xi(\kappa)|)^q] \\ & \leq & \mathbf{E}_f\left[\left(B^\beta(\kappa) + \sigma_\varepsilon(\gamma)|\xi(\kappa)|\right)^q\right] \\ & \leq & \mathbf{E}_f[(C^* S_\varepsilon(\kappa) + \sigma_\varepsilon(\gamma)|\xi(\kappa)|)^q] \\ & \leq & \left(\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}}\right)^q \mathbf{E}_f\left[\left(C^* + \frac{|\xi(\kappa)|}{\sqrt{\ln\ln\frac{1}{\varepsilon}}}\right)^q\right] \end{array}$$

b) Let us control $\mathbf{E}_f[I_2^q]$. Our procedure control itself this expectation. We have:

$$\mathbf{E}_f[I_2^q] \leq C^q \left(\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}}\right)^q.$$

Let us remark that we use the fact that κ belongs to $\mathcal{Z}_\gamma^\varepsilon$.
c) Finally, let us control $\mathbf{E}_f[I_1^q]$. Using lemma **??**, we ontain:

$$\begin{array}{rcl} \mathbf{E}_f[I_1^q] & = & \mathbf{E}_f\left[|\hat{f}_{\hat{k}}(t) - \hat{f}_{\hat{k}\wedge\kappa}|^q\right] \\ & \leq & \mathbf{E}_f\left[\left(2C^* S_\varepsilon(\kappa) + \sigma_\varepsilon(\hat{k})|\xi(\hat{k})| + \sigma_\varepsilon(\hat{k}\wedge\kappa)|\xi(\hat{k}\wedge\kappa)|\right)^q\right] \\ & \leq & S_\varepsilon(\kappa)^q \mathbf{E}_f\left[\left(2C^* + \frac{|\xi(\hat{k})| + |\xi(\hat{k}\wedge\kappa)|}{\sqrt{\ln\ln\frac{1}{\varepsilon}}}\right)^q\right] \\ & \leq & \mathbf{E}_f\left[\left(2C^* + \frac{|\xi(\hat{k})| + |\xi(\hat{k}\wedge\kappa)|}{\sqrt{\ln\ln\frac{1}{\varepsilon}}}\right)^q\right]\left(\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}}\right)^q \end{array}$$

Finally, we obtain the following inequelity:

$$\begin{array}{rcl} (*) & \leq & (3^{q-1} \vee 1)\left(\mathbf{E}_f[I_1^q] + \mathbf{E}_f[I_2^q] + \mathbf{E}_f[I_3^q]\right) \\ & \leq & (3^{q-1} \vee 1)\left\{C^q + (2^q + 1)(C^*)^q + o(1/\varepsilon)\right\}\left(\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}}\right)^q, \end{array}$$

where $o(1/\varepsilon)$ tends to 0 where ε tends to 0. It is clear by applying Lebesgue's theorem.
B) $\mathcal{A}$ is empty

As $\mathcal{A}$ is empty, in particular κ does not belong to this set. Thus, we obtain:

$$\begin{aligned}\mathbf{E}_f[|f_\varepsilon^\gamma(t)-f(t)|^q] &\leq \mathbf{E}_f\left[|f_\varepsilon^\Phi(t)-f(t)|^q\mathbf{1}_{\{\kappa\notin\mathcal{A}\}}\right]\\ &\leq \sqrt{\mathbf{E}_f[|f_\varepsilon^\Phi(t)-f(t)|^{2q}]\,\mathbf{P}_f[\kappa\notin\mathcal{A}]}\end{aligned}$$

Using the upper bound of the first part of this paper we obtain:

$$\sqrt{\mathbf{E}_f[|f_\varepsilon^\Phi(t)-f(t)|^{2q}]} \leq \mathrm{Cte}\left(\varepsilon\sqrt{\ln\frac{1}{\varepsilon}}\right)^{\frac{2\gamma}{2\gamma+1}q}.$$

Thus, we have to control $\mathbf{P}_f[\kappa\notin\mathcal{A}]$. If $\kappa\notin\mathcal{A}$, there exists $l\in\mathcal{Z}_\varepsilon^\gamma$, $l\neq\kappa$, such that:

$$|\hat{f}_{\kappa\wedge l}(t)-\hat{f}_l(t)| > C\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}}.$$

And, consequently, we obtain that:

$$\begin{aligned}\mathbf{P}_f[\kappa\notin\mathcal{A}] &\leq \sum_{l\neq\kappa}\mathbf{P}_f\left[|\hat{f}_{\kappa\wedge l}(t)-\hat{f}_l(t)| > C\sigma_\varepsilon(\gamma)\sqrt{\ln\ln\frac{1}{\varepsilon}}\right]\\ &\leq 2\sum_{\kappa\neq l}\left(\frac{1}{\ln\frac{1}{\varepsilon}}\right)^{\frac{(C-2C^*)^2}{8}}\\ &\leq 2(\#\mathcal{Z}_\gamma^\varepsilon)\left(\frac{1}{\ln\frac{1}{\varepsilon}}\right)^{\frac{(C-2C^*)^2}{8}}.\end{aligned}$$

Moreover, it is easy to prove that there exists a constant C_b depending only on b such that:

$$\#\mathcal{Z}_\gamma^\varepsilon \leq C_b\left(\ln\frac{1}{\varepsilon}\right)^d.$$

On the other hand our choice of C implies that

$$\frac{(C-2C^*)^2}{8} = d+\frac{2\gamma}{2\gamma+1}(2q).$$

Thus, we obtain:

$$\mathbf{E}_f[|f_\varepsilon^\gamma(t)-f(t)|^q] \leq \mathrm{Cte}\;\varepsilon^{\frac{2\gamma}{2\gamma+1}q}$$

4 Proof of (L.B.)

4.1 Method

The method is classical. Our goal is to minorate the minimax risk by a bayesian risk taken on a large number ($\sqrt{\ln 1/\varepsilon}$)of functions. In our mind, these functions are chosen because they represent the most difficult functions to be estimated in the considered class. This assertion is explained by lemma 2

4.2 Notations

Let us intoduce some basic notations. Let us fix $0 < \gamma < \bar{b}$. We say that a function $g : \mathbf{R}^d \to \mathbf{R}$ belongs to $\mathcal{G}(\gamma)$ if it satisfies:

$$\begin{cases} g(0) > 0. \\ \|g\| < +\infty \\ g \in \bigcap_{\beta \in \mathcal{B}(\gamma)} H(\beta) \\ \operatorname{supp} g \subset [-a; a]^d. \end{cases}$$

Here and later, we fix $g \in \mathcal{G}(\gamma)$.

Let us denote

$$\delta = \frac{\prod_{i=2}^d b_i}{\sum_{i=2}^d \prod_{j \neq i} b_j} = \frac{1}{1/\beta_2 + \cdots + 1/\beta_d}.$$

We consider

$$a = \left(\frac{1}{\gamma} - \frac{1}{\delta} \right)^{-1} < b_1,$$

and we denote $n_\varepsilon = \sqrt{\ln 1/\varepsilon}$.

Now, let us consider a family of vectors $\{\beta^{(k)}\}_k$ indexed by $k = 0, \ldots, n_\varepsilon$ and defined as follows:

$$\beta_1^{(k)} = a + k \frac{b_1 - a}{n_\varepsilon} \tag{1}$$

$$\beta_i^{(k)} = \frac{b_i}{\delta} \left(\frac{1}{\gamma} - \frac{1}{\beta_1^{(k)}} \right)^{-1} \quad \forall i = 2, \ldots, d. \tag{2}$$

Lemma 1. *For all $k = 0, \ldots, n_\varepsilon$ the vector $\beta^{(k)}$ belongs to $\mathcal{B}(\gamma)$.*

This lemma will be proved later.

Finally, let us introduce some functions. First of all, let us consider:

$$\forall i = 1, \dots, d,\ \forall k = 0, \dots, n_\varepsilon, \quad h_i^{(k)} = \left(\varkappa \varepsilon \sqrt{\ln \ln \frac{1}{\varepsilon}} \right)^{\frac{2\gamma}{2\gamma+1} \frac{1}{\beta_i^{(k)}}}$$

where $\varkappa < 1/(\sqrt{2}\|g\|)$. Then, we can define:

$$\left\{ \begin{array}{rcl} f_0 & \equiv & 0 \\ f_k(x) & = & \varkappa^{\frac{2\gamma}{2\gamma+1}} \eta_\varepsilon(\gamma) g \left(\frac{x_1 - t_1}{h_1^{(k)}}, \dots, \frac{x_d - t_d}{h_d^{(k)}} \right), \quad k \geq 1. \end{array} \right.$$

4.3 Proof

Now, let us prove our result. We will denote $\mathbf{P}_k$ instead of $\mathbf{P}_{f_k}$ and we consider the likelyhood ratio:

$$Z_\varepsilon = \frac{1}{n_\varepsilon} \sum_{k=1}^{n_\varepsilon} \frac{d\mathbf{P}_k}{d\mathbf{P}_0} (\mathcal{X}^{(\varepsilon)}).$$

This ratio satisfies the following lemma which will be proved further:

Lemma 2. *For all* $0 < \alpha < 1$*, we have:*

$$\limsup_{\varepsilon \to 0} \mathbf{P}_0[|Z_\varepsilon - 1| > \alpha] = 0.$$

Let us consider for any arbitrary estimator $\tilde{f}$, the following quantiy:

$$R_\varepsilon(\tilde{f}) = \sup_{f \in \Sigma_\gamma} \mathbf{E}_f \left[\left(\left(\varkappa \varepsilon \sqrt{\ln \ln \frac{1}{\varepsilon}} \right)^{-\frac{2\gamma}{2\gamma+1}} |\tilde{f}(t) - f(t)| \right)^q \right].$$

It is a well known result that, using bayesian method, for all $0 < \alpha < 1$ we obtain:

$$R_\varepsilon(\tilde{f}) \geq (1 - \alpha) \left(\frac{g(0)}{2} \right)^q (1 - \mathbf{P}_0[|Z_\varepsilon - 1| > \alpha]).$$

Thus, we have:

$$\liminf_{\varepsilon \to 0} R_\varepsilon(\tilde{f}) \geq (1 - \alpha) \left(\frac{g(0)}{2} \right)^q.$$

This inequality is equivalent to the following:

$$\liminf_{\varepsilon \to 0} \sup_{f \in \Sigma_\gamma} \mathbf{E}_f \left[\left(\eta_\varepsilon^{-1}(\gamma) |\tilde{f}(t) - f(t)| \right)^q \right] \geq (1 - \alpha) \left(\varkappa^{\frac{2\gamma}{2\gamma+1}} \frac{g(0)}{2} \right)^q.$$

Now, if $\varkappa$ tends to $(\sqrt{2}\|g\|)^{-1}$ and α tends to 1 we obtain the lower bound:

$$\liminf_{\varepsilon\to 0} \sup_{f\in\Sigma_\gamma} \mathbf{E}_f\left[\left(\eta_\varepsilon^{-1}(\gamma)|\tilde{f}(t)-f(t)|\right)^q\right] \geq \left(2^{-(1+\gamma/(2\gamma+1))} \sup_{g\in\mathcal{G}(\gamma)} \frac{g(0)}{\|g\|}\right)^q.$$

A Proof of lemma 1

First of all, let us prove that $a < b_1$. In fact:

$$a < b_1 \iff \frac{1}{b_1} < \frac{1}{\gamma} - \frac{1}{\delta}.$$

But it is clear that
$$\frac{1}{b_1} + \frac{1}{\delta} = \frac{1}{\bar{b}} < \frac{1}{\gamma}.$$

Result follows.

Let us fix $\beta \in \{\beta^{(k)}\}_k$.

Step 1. Let us calculate:

$$\begin{aligned} \sum_{i=1}^d \frac{1}{\beta_i} &= \frac{1}{\beta_1} + \sum_{i=2}^d \frac{\delta}{b_i}\left(\frac{1}{\gamma} - \frac{1}{\beta_1}\right) \\ &= \frac{1}{\gamma}. \end{aligned}$$

Step 2. Let us prove that, for all i, $\beta_i > 0$. First, we have $\beta_1 > a > 0$. Next, for $i \geq 2$, $\beta_i > 0$ if $1/\gamma > 1\beta_1$. But clearly we have $\beta_1 > a > \gamma$. Result follows.

Step 3. Let us prove that, for all i, $\beta_i \leq b_i$. This inequality is equivalent to:

$$\delta\left(\frac{1}{\gamma} - \frac{1}{\beta_1}\right) \geq 1,$$

i.e. $\beta_1 \geq a$. Finally, $\beta \in \mathcal{B}(\gamma)$. □

B Proof of lemma 2

First, let us remark that:

$$\mathbf{P}_0[|Z_\varepsilon - 1| > \alpha] \leq \alpha^{-2}\mathbf{E}_0\left[(Z_\varepsilon - 1)^2\right]$$

and, if $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbf{L}^2$,

$$\mathbf{E}_0\big[(Z_\varepsilon - 1)^2\big] = \frac{1}{n_\varepsilon^2} \sum_{k,l=1}^{n_\varepsilon} \exp\left(\frac{\langle f_k, f_l \rangle}{\varepsilon^2}\right) - 1.$$

It is enough to prove the following assertions:

$$\frac{1}{n_\varepsilon^2} \sum_{k=1}^{n_\varepsilon} \exp\left(\frac{\|f_k\|}{\varepsilon^2}\right) \xrightarrow[\varepsilon \to 0]{} 0, \tag{3}$$

and

$$\limsup_{\varepsilon \to 0} \frac{1}{n_\varepsilon^2} \sum_{k \neq l}^{n_\varepsilon} \exp\left(\frac{\langle f_k, f_l \rangle}{\varepsilon^2}\right) \leq 1 \tag{4}$$

First, let us prove Equation (3).

Let us calculate $\|f_k\|^2$ for all k. We have:

$$\|f_k\|^2 = \|g\|^2 \varkappa^2 \varepsilon^2 \ln \ln \frac{1}{\varepsilon} = 2\|g\|^2 \varkappa^2 \varepsilon^2 \ln n_\varepsilon.$$

Thus, we obtain:

$$\frac{1}{n_\varepsilon^2} \sum_{k=1}^{n_\varepsilon} \exp\left(\frac{\|f_k\|}{\varepsilon^2}\right) = n_\varepsilon^{2\|g\|^2 \varkappa^2 - 1}.$$

Thus, the choice of $\varkappa$ implies the result because $2\|g\|^2 \varkappa^2 - 1 < 0$.

Now, let us prove Equation (4).

Let us fix $1 \leq k < l \leq n_\varepsilon$. By an easy computation we obtain:

$$\langle f_k, f_l \rangle \leq \varkappa^{\frac{4\gamma}{2\gamma+1}} \eta_\varepsilon^2(\gamma) \|g\|_\infty^2 \mathrm{Vol}(C_k \cap C_l),$$

where Vol is the standard volume in $\mathbf{R}^d$ and C_k denotes the support of f_k:

$$C_k = \prod_{i=1}^{d} [-a h_i^{(k)}; a h_i^{(k)}].$$

Clearly, $h_1^{(k)} < h_1^{(l)}$ and, for any $i \geq 2$, we have $h_i^{(k)} > h_i^{(l)}$. Thus,w e can conclude that:

$$\mathrm{Vol}(C_k \cap C_l) = (2a)^d \frac{h_1^{(k)}}{h_1^{(l)}} \left(\prod_{i=1}^{d} h_i^{(l)}\right) \leq (2a)^d \frac{h_1^{(k)}}{h_1^{(k+1)}} \left(\prod_{i=1}^{d} h_i^{(l)}\right).$$

Let us calculate $h_1^{(k)}/h_1^{(k+1)}$:

$$\begin{aligned}
\frac{h_1^{(k)}}{h_1^{(k+1)}} &= \left(\varkappa^{\frac{2\gamma}{2\gamma+1}}\eta_\varepsilon(\gamma)\right)^{1/\beta_1^{(k)}-1/\beta_1^{(k+1)}} \\
&= \left(\varkappa^{\frac{2\gamma}{2\gamma+1}}\eta_\varepsilon(\gamma)\right)^{\frac{1/n_\varepsilon}{\beta_1^{(k)}\beta_1^{(k+1)}}} \\
&\le \left(\varkappa^{\frac{2\gamma}{2\gamma+1}}\eta_\varepsilon(\gamma)\right)^{\frac{1}{b_1^2 n_\varepsilon}}.
\end{aligned}$$

Moreover, let us remark that:

$$\prod_{i=1}^{d} h_i^{(l)} = \varkappa^{\frac{2}{2\gamma+1}}\eta_\varepsilon^{1/\gamma}(\gamma).$$

Then, by an easy computation, we deduce that:

$$\langle f_k, f_l\rangle \le (2a)^d(\varkappa^{1+\frac{\Gamma}{n_\varepsilon}}\|g\|_\infty)^2\left(\eta_\varepsilon(\gamma)\right)^{\frac{2\gamma+1}{\gamma}\left(1+\frac{\Gamma}{n_\varepsilon}\right)},$$

where

$$\Gamma = \frac{\gamma}{b_1^2(2\gamma+1)}.$$

Let us recall that $\eta_\varepsilon(\gamma) = (\varepsilon\sqrt{\ln\ln 1/\varepsilon})^{2\gamma/(2\gamma+1)}$. Thus we obtain:

$$\frac{\langle f_k, f_l\rangle}{\varepsilon^2} \le (2a)^d(\varkappa^{1+\frac{\Gamma}{n_\varepsilon}}\|g\|_\infty)^2\mathcal{M}_\varepsilon,$$

where

$$\mathcal{M}_\varepsilon = \left(\ln\ln\frac{1}{\varepsilon}\right)\left(\varepsilon^2\ln\ln\frac{1}{\varepsilon}\right)^{\frac{\Gamma}{n_\varepsilon}}$$

teds to 0 when ε tends to 0 (it is easy to see that $\ln\mathcal{M}_\varepsilon \to -\infty$).

Now, let us back to Equation (4):

$$\frac{1}{n_\varepsilon^2}\sum_{k\neq l}^{n_\varepsilon}\exp\left(\frac{\langle f_k, f_l\rangle}{\varepsilon^2}\right) \le \frac{n_\varepsilon - 1}{n_\varepsilon}\exp\left((2a)^d(\varkappa^{1+\frac{\Gamma}{n_\varepsilon}}\|g\|_\infty)^2\mathcal{M}_\varepsilon\right) \xrightarrow[\varepsilon\to 0]{} 1.$$

And Lemma is proved. □

References

[1] Andrew Barron, Lucien Birgé, and Pascal Massart. Risk bounds for model selection via penalization. *Probab. Theory Related Fields*, 113(3):301–413, 1999.

[2] E. N. Belitser and B. Y. Levit. On minimax filtering over ellipsoids. *Math. Methods Statist.*, 4(3):259–273, 1995.

[3] E. N. Belitser and B. Y. Levit. Asymptotically minimax nonparametric regression in L_2. *Statistics*, 28(2):105–122, 1996.

[4] Karine Bertin. Asymptotically exact minimax estimation in sup-norm for anisotropic Hölder classes. *Bernoulli*, 10(5):873–888, 2004.

[5] Karine Bertin. Minimax exact constant in sup-norm for nonparametric regression with random design. *J. Statist. Plann. Inference*, 123(2):225–242, 2004.

[6] Lucien Birgé. Approximation dans les espaces métriques et théorie de l'estimation. *Z. Wahrsch. Verw. Gebiete*, 65(2):181–237, 1983.

[7] Lucien Birgé. Nonasymptotic minimax risk for Hellinger balls. *Probab. Math. Statist.*, 5(1):21–29, 1985.

[8] Lucien Birgé and Pascal Massart. From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York, 1997.

[9] Lucien Birgé and Pascal Massart. Gaussian model selection. *J. Eur. Math. Soc. (JEMS)*, 3(3):203–268, 2001.

[10] L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3):843–874, 2002. Dedicated to the memory of Lucien Le Cam.

[11] L. Cavalier, Y. Golubev, O. Lepski, and A. Tsybakov. Block thresholding and sharp adaptive estimation in severely ill-posed inverse problems. *Teor. Veroyatnost. i Primenen.*, 48(3):534–556, 2003.

[12] L. Cavalier and A. B. Tsybakov. Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation. *Math. Methods Statist.*, 10(3):247–282, 2001. Meeting on Mathematical Statistics (Marseille, 2000).

[13] Albert Cohen, Ronald DeVore, Gerard Kerkyacharian, and Dominique Picard. Maximal spaces with given rate of convergence for thresholding algorithms. *Appl. Comput. Harmon. Anal.*, 11(2):167–191, 2001.

[14] David L. Donoho and Iain M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.

[15] David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.*, 90(432):1200–1224, 1995.

[16] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: asymptopia? *J. Roy. Statist. Soc. Ser. B*, 57(2):301–369, 1995. With discussion and a reply by the authors.

[17] David L. Donoho, Iain M. Johnstone, Gérard Kerkyacharian, and Dominique Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2):508–539, 1996.

[18] S. Yu. Efroĭmovich and M. S. Pinsker. Estimation of square-integrable probability density of a random variable. *Problemy Peredachi Informatsii*, 18(3):19–38, 1982.

[19] Sam Efromovich. *Nonparametric curve estimation*. Springer Series in Statistics. Springer-Verlag, New York, 1999. Methods, theory, and applications.

[20] V. N. Gabušin. Best approximation of functionals on certain sets. *Mat. Zametki*, 8:551–562, 1970.

[21] A. Goldenshluger and A. Nemirovski. On spatially adaptive estimation of nonparametric regression. *Math. Methods Statist.*, 6(2):135–170, 1997.

[22] G. K. Golubev. Adaptive asymptotically minimax estimates for smooth signals. *Problemy Peredachi Informatsii*, 23(1):57–67, 1987.

[23] Y. Golubev and B. Levit. An oracle approach to adaptive estimation of linear functionals in a Gaussian model. *Math. Methods Statist.*, 13(4):392–408 (2005), 2004.

[24] Yu. Golubev, O. Lepski, and B. Levit. On adaptive estimation for the sup-norm losses. *Math. Methods Statist.*, 10(1):23–37, 2001.

[25] Wolfgang Härdle, Gerard Kerkyacharian, Dominique Picard, and Alexander Tsybakov. *Wavelets, approximation, and statistical applications*, volume 129 of *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.

[26] R. Z. Hasminskii and I. A. Ibragimov. On asymptotic efficiency in the presence of an infinite-dimensional nuisance parameter. In *Probability theory and mathematical statistics (Tbilisi, 1982)*, volume 1021 of *Lecture Notes in Math.*, pages 195–229. Springer, Berlin, 1983.

[27] I. A. Ibragimov and R. Z. Hasminski. On the estimation of a signal, its derivatives and the maximum point for Gaussian observations. *Teor. Veroyatnost. i Primenen.*, 25(4):718–733, 1980.

[28] I. A. Ibragimov and R. Z. Hasminski. Some problems of nonparametric estimation of the value of a linear functional. *Dokl. Akad. Nauk SSSR*, 256(4):781–783, 1981.

[29] I. A. Ibragimov and R. Z. Hasminski. *Statistical estimation*, volume 16 of *Applications of Mathematics*. Springer-Verlag, New York, 1981. Asymptotic theory, Translated from the Russian by Samuel Kotz.

[30] Ildar Abdulovic Ibragimov and Rafail Zalmanovic Hasminski. Asymptotic properties of some nonparametric estimates in Gaussian white noise. In *Third International Summer School on Probability Theory and Mathematical Statistics (Varna, 1978) (Bulgarian)*, pages 29–64. B lgar. Akad. Nauk, Sofia, 1980.

[31] Gérard Kerkyacharian, Oleg Lepski, and Dominique Picard. Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields*, 121(2):137–170, 2001.

[32] Gérard Kerkyacharian and Dominique Picard. Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.*, 24(2):485–507, 1996.

[33] Gérard Kerkyacharian and Dominique Picard. Minimax or maxisets? *Bernoulli*, 8(2):219–253, 2002.

[34] Gérard Kerkyacharian, Dominique Picard, and Karine Tribouley. L^p adaptive density estimation. *Bernoulli*, 2(3):229–247, 1996.

[35] A. P. Korostelev. An asymptotically minimax regression estimator in the uniform norm up to a constant. *Teor. Veroyatnost. i Primenen.*, 38(4):875–882, 1993.

[36] A. P. Korostelev and A. B. Tsybakov. Asymptotically minimax image reconstruction problems. In *Topics in nonparametric estimation*, volume 12 of *Adv. Soviet Math.*, pages 45–86. Amer. Math. Soc., Providence, RI, 1992.

[37] Alexander Korostelev and Michael Nussbaum. The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. *Bernoulli*, 5(6):1099–1118, 1999.

[38] O. V. Lepski. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990.

[39] O. V. Lepski. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 36(4):645–659, 1991.

[40] O. V. Lepski and B. Y. Levit. Adaptive minimax estimation of infinitely differentiable functions. *Math. Methods Statist.*, 7(2):123–156, 1998.

[41] O. V. Lepski and B. Y. Levit. Adaptive nonparametric estimation of smooth multivariate functions. *Math. Methods Statist.*, 8(3):344–370, 1999.

[42] O. V. Lepski, E. Mammen, and V. G. Spokoiny. Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.*, 25(3):929–947, 1997.

[43] O. V. Lepski and V. G. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, 25(6):2512–2546, 1997.

[44] O. V. Lepski and A. B. Tsybakov. Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probab. Theory Related Fields*, 117(1):17–48, 2000.

[45] O. V. Lepskiĭ. On problems of adaptive estimation in white Gaussian noise. In *Topics in nonparametric estimation*, volume 12 of *Adv. Soviet Math.*, pages 87–106. Amer. Math. Soc., Providence, RI, 1992.

[46] E. Mammen. On qualitative smoothness of kernel density estimates. *Statistics*, 26(3):253–267, 1995.

[47] Enno Mammen. A short note on optimal bandwidth selection for kernel estimators. *Statist. Probab. Lett.*, 9(1):23–25, 1990.

[48] Michael Nussbaum. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, 24(6):2399–2430, 1996.

[49] Valentin V. Petrov. *Limit theorems of probability theory*, volume 4 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, New York, 1995. Sequences of independent random variables, Oxford Science Publications.

[50] Dominique Picard and Karine Tribouley. Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.*, 28(1):298–335, 2000.

[51] Vincent Rivoirard. Maxisets for linear procedures. *Statist. Probab. Lett.*, 67(3):267–275, 2004.

[52] Vincent Rivoirard. Thresholding procedure with priors based on Pareto distributions. *Test*, 13(1):213–246, 2004.

[53] A. V. Skorohod. *Integration in Hilbert space*. Springer-Verlag, New York, 1974. Translated from the Russian by Kenneth Wickwire, Ergebnisse der Mathematik und ihrer Grenzgebiete, Band 79.

[54] Charles J. Stone. Optimal rates of convergence for nonparametric estimators. *Ann. Statist.*, 8(6):1348–1360, 1980.

[55] Charles J. Stone. Admissibility and local asymptotic admissibility of procedures which combine estimation and model selection. In *Statistical decision theory and related topics, III, Vol. 2 (West Lafayette, Ind., 1981)*, pages 317–333. Academic Press, New York, 1982.

[56] Charles J. Stone. An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, 12(4):1285–1297, 1984.

[57] Charles J. Stone. Additive regression and other nonparametric models. *Ann. Statist.*, 13(2):689–705, 1985.

[58] Hans Triebel. *Theory of function spaces. II*, volume 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1992.

[59] A. B. Tsybakov. Asymptotically efficient estimation of a signal in L_2 under general loss functions. *Problemy Peredachi Informatsii*, 33(1):94–106, 1997.

[60] A. B. Tsybakov. Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26(6):2420–2469, 1998.

[61] Alexandre B. Tsybakov. *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 2004.